DOCUMENT RESUME

ED 457 440 CG 031 176

AUTHOR Vacc, Nicholas A.; Loesch, Larry C.; Lubik, Ruth E.
TITLE Writing Multiple-Choice Test Items.
PUB DATE 2001-00-00
NOTE 10p.; In its: Assessment: Issues and Challenges for the Millennium; see CG 031 161.
PUB TYPE Reports - Descriptive (141)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Assessment; *Evaluation; Item Analysis; *Measures (Individuals); Multiple Choice Tests; Questioning Techniques; Testing Problems; *Tests

ABSTRACT

Multiple choice tests are widely viewed as the most effective and objective means of assessment. Item development is the central component of creating an effective test, but test developers often do not have the background in item development. This document describes recall, application, and analysis, the three cognitive levels of test items. It also reviews single best response, situational set, and complex, the three main item types. A discussion is provided of guidelines for writing appropriate and effective item stems, keyed responses, and distracters. (Contains 11 references.) (JDM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 457 440

Writing Multiple-Choice Test Items

By Nicholas A. Vacc, Larry C. Loesch, & Ruth E. Lubik

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

CG 031 176

BEST COPY AVAILABLE

Chapter Nineteen

Writing Multiple-Choice Test Items

Nicholas A. Vacc, Larry C. Loesch, & Ruth E. Lubik

Abstract

Multiple-choice tests are widely viewed as the most efficient and objective means of assessment. Item development is the most critical component of creating an effective test, but unfortunately, most test developers have no background in item development. The three cognitive levels of test items (recall, application, and analysis) are described, along with the three main item types (single best response, situational set, and complex). Finally, guidelines for writing appropriate and effective item stems, keyed responses, and distracters are provided.

Most adults have taken a multiple-choice test at some time in their lives. Such tests frequently are used in educational systems to assess academic aptitude or achievement, and they frequently are used in job application processes to determine an applicant's potential or skills. They also often are used in professions as part of a licensure or certification application process (Karras, 1991; Vacc, 1991). Clearly, tests are viewed by many as the best and most efficient way to gather and evaluate data and information.

Because multiple-choice tests are used widely and because they have significant impact on the lives of those taking them, using procedures that are proven effective for their development is important. Cohen and Swerdlik (1999, p. 215) indicated, "The creation of a good test is not a matter of chance—it is the product of the thoughtful and sound application of established principles of test construction." Such principles are found in resources such as the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1985),

Responsibilities of Users of Standardized Tests (AACD & AMECD, 1989), and *Code of Fair Testing Practices* (Joint Committee on Testing Practices, 1988). Each set of principles has as its goal the development of an instrument that has a high level of objectivity and validity, because well-produced tests increase the likelihood that test scores can be of assistance (Vacc, 1991).

Haladyna and Downing (1989) noted that one of the most important steps in test development is item writing. They concluded that test quality therefore is contingent upon the quality of test items. Unfortunately, McDougall (1997) and Osterlind (1989) stated that most test developers construct tests based on "folk wisdom" rather than a systematic application of principles of effective item development. Most likely, the lack of a systematic procedure occurs because few professionals are trained adequately in test construction; therefore, they focus on test information interesting to themselves rather than on essential material. The unfortunate result often is item-writer bias (Haladyna, 1992; McDougall, 1997). Even highly educated college faculty typically lack effective test-development training and thus make similar errors (McDougall, 1997).

Despite common and widespread problems in test construction, multiple-choice tests remain popular and appear to be dominant among objective tests (Haladyna, 1992; Haladyna & Downing, 1989; McDougall, 1997; Pomplun & Omar, 1997). Multiple-choice tests afford fast, relatively accurate, economical, and objective ways to obtain data, and they have the advantage of being applicable to a wide range of topics (Cohen & Swerdlik, 1999). Multiple-choice tests also are generally thought to be reliable, versatile, and easily used (Haladyna & Downing, 1989; Karras, 1991; McDougall, 1997).

Haladyna (1992) suggested that better measurement of both achievement and abilities could be achieved most easily through improvements in item writing. Haladyna and Downing (1989, p. 47) compiled 43 item-writing guidelines, rules, and suggestions from various textbooks, and concluded that applying these guidelines would result in tests that are uniform in appearance and free of nettlesome item-writing faults and other problems that distract examinees from giving their best responses.

Most multiple-choice items can be classified into one of three cognition levels: recall, application, and analysis. Each level utilizes a different cognitive function:

Recall-level items: Recall-level items primarily test the recognition or recall of relatively isolated facts, concepts, principles, processes, procedures, or theories. Responding correctly to items at this level is primarily a function of an individual's memory. Incorrect

responses result when the individual is unable to remember or recall the answer.

Application-level items: Application-level items primarily test relatively simple interpretations or limited applications of data or information. Items at this level require more than application of memory; responding correctly requires relatively minor or low-level problem-solving skills.

Analysis-level items: The third commonly used level of multiple-choice items is the analysis level. Items at this level primarily test skills involving evaluation of data, problem solving, or the fitting together of elements into a meaningful whole. Responding correctly to these items involves application of both good judgment and problem-solving skills. This level thus involves higher cognitive processes than the other levels.

Item Types

Multiple-choice items also can be classified by type, with each type having unique characteristics and challenging a respondent's thinking in different ways. Three commonly used types of multiple-choice items are single best response, situational set, and complex.

Single best response items: The most commonly used type is the single best response item. With this type of item, there purportedly is one correct answer among the various response choices (sometimes called the distracters or foils) for the item. Single best response items may be developed in several forms. One form is the direct question in the item stem to which the respondent is required to provide the answer from the response choices. Another form is an incomplete statement in the item stem for which the respondent is asked to select the word or phrase from among the choices that best completes it. The third form is the calculation item for which the respondent is required to perform some calculation, usually mathematical, in order to determine the correct response from among the choices.

Situational set items: The situational set item presents a scenario containing a collection of facts or data, followed by the item stem. Typically, there are three to five multiple choices associated with each situational set, usually of the single best response form. However, each choice is expected to stand alone and is not contingent upon any other for correct responding.

Complex items: The complex item requires simultaneous consideration of several facts or bits of information. A complex item consists of a stem followed by three to five statements, phrases, or sometimes graphic depictions known as the elements. The distracters in the item include combinations of the elements. Respondents to these types of items face an all-or-none dilemma; knowing only one of the elements will not allow determination of the correct response.

Writing Multiple-Choice Items

Theoretically, the correct way to respond to a multiple-choice test question is not by eliminating the incorrect responses and then choosing from the remaining responses, but rather by reading the item stem carefully, formulating the correct response based on the information in the item stem, and then finding the correct response from among the distracters. The approach to responding has significant implications for writing effective multiple-choice items. For example, the item stem must be written so that respondents can formulate the correct response mentally before considering the distracters. In addition, effective distracters are created through consideration of how respondents might think incorrectly or illogically in responding to the item stem.

Writing Item Stems

There are several guidelines to follow in constructing item stems effectively and efficiently. One is to use clear and simple language. The use of jargon and highly technical vocabulary should be avoided unless they are appropriate for the purpose of the item. An item developer also should use simple sentences and grammatical constructions that promote ease of reading and understanding for the respondent.

A second guideline in stem construction is to present only a single, clearly formulated idea or problem. Item developers should avoid including multiple ideas or vague or ambiguous concepts in the item stem. In addition, test items should focus on general knowledge and principles and be devoid of unnecessary specificity; excessive "window dressing" or irrelevant information defeats the goal of effective assessment.

The last major item-stem development guideline is to put as much of the wording as possible in the stem rather than writing a short item stem with numerous distracters. In fact, all the information or qualifications necessary to determine the correct answer should be in the item stem. At the same time, however, item developers should avoid using a literal definition as the item stem. Rather, the stem should

provide the information in clear, easily understood language. Finally, the use of negative wording (e.g., "which of the following is not") should be avoided as much as possible.

Writing Distracters

Formulating distracters with care is important so that irrelevant characteristics do not trigger responding behaviors. Foremost, an item developer must insure that the keyed response (i.e., the one to be scored as correct) is both correct and clearly the best response. The distracters in a multiple-choice item should be independent of one another, arranged in logical order, and grammatically consistent with the stem. They also should not cue responding to answers or distracters in other items. In general, item developers should avoid using phrases such as "all of the above" or "none of the above" as distracters.

Multiple-choice item distracters should be designed to be attractive to respondents who do not have a good understanding of the content of the item stem. One reasonably effective method of constructing such distracters is to use common misconceptions about the content in the item stem. Using "good-sounding" words in the distracters, such as *accurate, important,* or *significant* often is effective. Also, good distracters should be similar to the keyed response in length, complexity, and grammatical structure. Presenting distracters in language familiar to respondents and avoiding distracters that contradict each other are other effective strategies.

General Guidelines for Test Items

A test developer must decide upon the most effective and efficient format possible for testing the desired material. Irrelevant sources of difficulty should be avoided, as should items that cue responses for other items. Normal and correct rules of grammar and spelling should be used and the use of gender-specific pronouns should be avoided.

If the stem is a question, each distracter should begin with a capital letter and end with a period because the distracters are not continuations of the item stem. When the item stem is an incomplete sentence, each distracter should begin with a lower-case letter. Periods should be omitted following numeric distracters to avoid confusion with decimal points.

Irrelevant clues to the keyed response should be avoided by having essentially similar language in the stem and the keyed response and by avoiding buzzwords that give away the keyed response. Additionally, vague modifiers, such as *sometimes, usually,* or *may,* should be avoided, as should absolute terms such as *always, never, none,* or *only.* Essentially equivalent distracters should also be avoided.

Other important concerns in effective item development are to keep the reading level of the item stem and distracters as low as possible, and to avoid the repetitive use of favorite phrases, terms, or grammatical constructions. Items or questions for which the correct response is merely an opinion also should be avoided unless the source of the opinion is identified clearly. Item content tied to a specific reference, such as a textbook or journal article, should be avoided unless a particular perspective is being espoused, in which case the source must be identified clearly.

It is good psychometric practice to have items reviewed for clarity and cogency before their initial administration, preferably by persons similar to the intended respondents. Item performance characteristics also need to be examined after each administration, particularly those relative to item difficulty, discrimination, reliability, and validity. In effect, each item is field tested in each administration by reviewing the results and item data, and revising as appropriate.

Conclusion

Knowing how to construct good multiple-choice items has important implications for counselors. Indeed, the codes of ethics of the American Counseling Association and the National Board for Certified Counselors call for professional counselors to be knowledgeable of testing and test construction. These admonitions are made because counselors frequently are involved in test use and evaluation, either as test users or test developers, and they frequently help develop tests that are used to evaluate other individuals. In addition, important and significant judgments about individuals and programs are made based on test scores. Thus, if counselors are to fulfill their professional functions and obligations effectively and fully, they must be knowledgeable in effective test- and item-development practices.

References

American Association for Counseling and Development (AACD) & Association for Measurement and Evaluation in Counseling and Development (AMECD). (1989). *Standards for users of standardized tests*. Alexandria, VA: AACD.

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). (1985). *Standards for educational and psychological testing*. Washington, DC: APA.

Cohen, R. J., & Swerdlik, M. E. (1999). *Psychological testing and assessment: An introduction to tests and measurement* (4th ed.). Mountain View, CA: Mayfield.

Haladyna, T. M. (1992). The effectiveness of several multiple-choice formats. *Applied Measurement in Education, 5*, 73–88.

Haladyna, T. M., & Downing, S. M. (1989). A taxonomy of multiple-choice item writing rules. *Applied Measurement in Education, 2*, 37–50.

Joint Committee on Testing Practices. (1988). *Code of fair testing practices*. Washington, DC: APA.

Karras, R. W. (1991). Let's improve multiple-choice tests. *OAH Magazine of History, 6*, 8–9.

McDougall, D. (1997). College faculty's use of objective tests: State-of-the practice versus state-of-the-art. *Journal of Research and Development in Education, 30*, 183–193.

Osterlind, S. J. (1989). *Constructing test items*. Boston, MA: Kluwer Academic.

Pomplun, M., & Omar, M. H. (1997). Multiple-mark items: An alternative objective item format. *Educational and Psychological Measurement, 57*, 949–962.

Vacc, N. A. (1991). Changing times: Changing views about testing. *Journal of Humanistic Education and Development, 30*, 88–90

About the Authors

Nicholas A. Vacc is the Joe Rosenthal Excellence Professor and past chairperson of the Department of Counseling and Educational Development at the University of North Carolina at Greensboro. In 1991 and 1994, during his tenure as chair, the department was selected as the Outstanding Counselor Education Program by the Association for Counselor Education and Supervision. He has served as president of the Association for Assessment in Counseling, editor of *Measurement and Evaluation in Counseling and Development*, and president of Chi Sigma Iota International. He also is a member of numerous technical

and policy advisory boards and working groups concerned with assessment in counseling and mental health, including the Joint Committee on Testing Practices.

Larry C. Loesch is currently a professor in the Department of Counselor Education at the University of Florida, where he has taught since he received his Ph.D. from Kent State University in 1973. A national certified counselor, his teaching interests include research, measurement and evaluation, and supervision of students' research and counseling. Loesch served as president of both the Florida and national Associations for Measurement and Evaluation in Guidance, the Florida Counseling Association, SACES, and Chi Sigma Iota. He was a co-recipient of the American Counseling Association's 1983 Research Award and of its 1992 Arthur A. Hitchcock Distinguished Professional Service Award. Loesch was a Fulbright Scholar in Slovakia in 2001.

Ruth E. Lubik is a substance abuse counselor in the Randolph County, North Carolina Mental Health Department. She is a graduate of the University of North Carolina at Greensboro with a M.S. and an Ed.S. from the Counseling and Educational Development Department.